Supporting Information

Structural Evidence for Asymmetrical Nucleotide Interactions in Nitrogenase

F. Akif Tezcan[1]*, Jens T. Kaiser[2], James B. Howard[3] and Douglas C. Rees[2]*

[1]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093-0356; [2]Division of Chemistry and Chemical Engineering and the Howard Hughes Medical Institute, California Institute of Technology, Pasadena, CA 91125; [3]Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN 55455.

Materials and Methods

Protein preparation and crystallization: Fe-protein and MoFe-protein from *Azotobacter vinelandii* were isolated and purified under anaerobic conditions as described previously.[1] Crystals of the *pcp/adp*-complex were prepared under conditions similar to the *pcp*-complex[2] in sitting drops by vapor diffusion in an anaerobic chamber at room temperature, using solutions and chemicals that were thoroughly deaerated using vacuum/Ar-fill cycles. Precipitating solutions contained 18-22% PEG 8000 (w/v), in addition to 0-50 mM NaCl, 100 mM Tris buffer (pH 8.5), and 10 mM sodium dithionite. The protein solutions were prepared by mixing appropriate amounts of Fe-protein and MoFe-protein stock solutions ([Fe-protein]stock= ~40 mg/ml, [MoFe-protein]stock= ~20 mg/ml), which were in 100 mM Tris (pH 7.75), 200 mM NaCl and 5 mM dithionite. Complex cocrystals were obtained using 1:1.2 to 1:1.5 MoFe-protein:Fe-protein ratios (v/v), corresponding to ~4 to 6-fold molar excess of Fe-protein per active site of MoFe-protein. The protein solutions also contained 10 mM $MgCl_2$, 5 mM AMPPCP and 5 mM ADP, whose stock solutions (100 mM) were prepared in 100 mM Tris buffer (pH 8.5) immediately prior to setting up the crystallization trays. The crystallization drops

contained 2 µl of the precipitating solution and 2 µl of the protein mixture, and the reservoir contained 250-500 µl of the precipitating solution.

Crystallography: For the collection of X-ray diffraction data, suitable crystals were exchanged into 25% PEG 400 in five steps over one hour for cryoprotection and flash-frozen in liquid nitrogen. X-ray diffraction data were collected at ALS (BL 8.2.1) using 1.0-Å radiation. Data collection and refinement statistics are summarized in Table S1. Data sets were processed using XDS[3] and AIMLESS.[4] The initial phases were obtained by molecular replacement using MOLREP[5] with MoFe-protein (PDB ID 1M1N) as a search model. Initial electron density maps clearly revealed the positions of the iron-sulfur clusters and most secondary structure elements of Fe-protein molecules, allowing their manual placement using the ADP•AlF_4^- complexed Fe-protein (PDB ID 1M34)[6] as the model. After an initial round of rigid-body refinement with CNS[7] and manual rebuilding with MAIN,[8] the model was refined with PHENIX[9] with alternating rounds of manual rebuilding in MAIN and COOT.[10] The stereochemistry of the final model was calculated using MOLPROBITY[11] as implemented in PHENIX. Final refinement statistics are given in Table S1. For consistency, the *pdp*-complex structure[2] (PDB ID 2AFK) was re-refined against the original diffraction data using the same protocol, with final refinement statistics listed in Table S2. Figures were prepared using PYMOL.[12]

Simulated annealing omit maps were calculated by phenix.refine (Fig S1), using the final coordinate file with all nucleotides and magnesium-water molecules deleted, applying the default values for cartesian simulated annealing and automated torsion NCS. Final R/R-free values were 0.153/0.195.

Principal Components Analysis

A principal components analysis (PCA) of the conformational changes relating a set of 21 Fe-protein coordinates (Table S3) was conducted using methods developed by Berendsen and coworkers.[13,14] For these calculations, a reference Fe-protein dimer (chain IDs E and F of PDB ID 1N2C) was transformed so that the molecular two-fold axis was positioned on the z axis of an orthogonal coordinate frame (by convention, the chain IDs of the two Fe-protein subunits in a dimer are (A, B) for free Fe-protein and (E, F), (G, H), (M, N) and (O, P) when complexed with MoFe-protein, where E, G, M, O primarily interact with the alpha-subunit). One subunit (chain ID equal to A, E, G, M or O) from each of the 20 other Fe-protein dimers was then superimposed onto chain E of the reference Fe-protein dimer. The PCA was then used to identify the directions associated with the largest concerted movement of atoms (ie, the principal components) that best describe the relationships between the second subunits (ie corresponding to chain IDs B, F, H, N or P). The α-carbons of residues 6-49, 70-115, 120-125, 129-186 and 193-260 (222/289 residues) were used for the superposition; residues in flexible or more highly variable regions were excluded to emphasize the rigid body type motions.

The basic algorithm involves the calculation of the covariance matrix of the fluctuations of each atomic coordinate from the average value of the corresponding coordinate in Fe-protein subunits with chain IDs B, F, H, N or P:

$$C_{ij} = \sum_{i=1}^{3N_{res}} \sum_{j=1}^{3N_{res}} \sum_{k=1}^{N_{mol}} \left(x_{i,k} - x_{i,ave}\right)\left(x_{j,k} - x_{j,ave}\right)$$

The eigenvectors of this matrix are directions in a $3N_{res}$-dimensional space (where N_{res} is the number of residues = 222 (the factor of 3 accounts for the individual x, y and z components) and N_{mol} is the number of molecules = 21); motions along an eigenvector represent concerted fluctuations of residues. The eigenvalues are the sum of the squared deviations of the system for each eigenvector. The value of an eigenvalue relative to the total squared displacement of all structures from the average indicates the proportion of the total deviation contributed by that eigenvector.

The displacement of the k^{th} structure along the j^{th} eigenvector is given by the projection:

$$q_{j,k} = \sum_{i=1}^{3N_{res}} \left(x_{i,k} - x_{i,ave} \right) \times \text{eigvec}(i,j)$$

$$\left| q_{j,k} \right| = rms \times \sqrt{N_{res}}$$

To calculate the rigid body rotation/transformation corresponding to the j^{th} eigenvector, the displacement along that eigenvector corresponding to the structure k with the largest value of $q_{j,k}$ is calculated; this displacement is added to the reference structure and the rigid body transformation calculated by an algorithm described by Kabsch.[15]

The components along the top 3 eigenvectors calculated for the PCA of Fe-protein conformations are tabulated in Table S3. The rotation axes corresponding to eigenvectors 1 and 2 (EV1 and EV2) are depicted in text Figure 3. The sum of the squared displacements of all 222 residues in the 21 structures from the average of all their positions is 82,441 $Å^2$; the projections along EV1 and EV2 account for 72,617 $Å^2$ (88%) and 5,430 $Å^2$ (~7%), respectively, so that the contribution from all the remaining eigenvectors is ~5%.

EV1 corresponds to a hinge bending motion as the two Fe-protein subunits qualitatively move approximately perpendicular to the dimer interface, while EV2 corresponds to a twisting motion as the Fe-protein subunits move approximately parallel to the dimer interface relative to one another. For this calculation, a translation of 2 Å along an eigenvector corresponds to ~5˚ rotation about the corresponding rotation axis. The hinge angle is the rotation angle required to relate F/B/N/P subunits to that of 1N2C subunit F (the original *alf*-complex structure) when E/A/M/O subunits are all superimposed on subunit E of 1N2C. Consequently, the reference state is the *alf*-complex and the hinge rotation is defined relative to that state. In Figure 3, the distance between the *alf*-complexes and the nucleotide free Fe-protein is ~10 Å along component 1, which corresponds to ~25˚ rotation to relate these two conformations, as is observed for the hinge angle.

Table S1. Data collection and refinement statistics for the *pcp/adp*-complex[*]

PDB Accession Code	4WZA
Wavelength (Å)	1.000
Resolution range (Å)	19.8 - 1.9 (1.967 - 1.9)
Space group	P 2_1 2_1 2_1
Unit cell (a, b, c (Å))	110.201, 120.412, 264.318
Total reflections	973,682 (69,288)
Unique reflections	269,423 (24,317)
Multiplicity	3.6 (2.8)
Completeness (%)	97.8 (87.8)
Mean I/sigma(I)	13.6 (2.1)
Wilson B-factor	19.80
R-merge	0.078 (0.550)
R-work	0.145 (0.236)
R-free	0.186 (0.281)
Number of non-hydrogen atoms	26,664
macromolecules	24,234
ligands	246
water	2,184
Protein residues	3,110
rms bonds (Å)	0.010
rms angles (°)	1.51
Ramachandran outliers (%)	0.36
Clashscore	2.31
Average B-factor (Å^2)	28.9
macromolecules	28.4
ligands	24.2
solvent	34.9

[*]Statistics for the highest-resolution shell are shown in parentheses.

Table S2. Refinement statistics for the re-refined 2AFK *pcp*-complex*

PDB Accession Code	4WZB
Resolution range (Å)	49.38 - 2.2 (2.279 - 2.2)
Space group	P 2_1 2_1 2_1
Unit cell (a, b, c (Å))	110.533, 120.894, 264.834
Reflections used for R-free	
R-work	0.193 (0.243)
R-free	0.254 (0.312)
Number of non-hydrogen atoms	25,159
macromolecules	23,976
ligands	240
water	943
Protein residues	3,064
rms bonds (Å)	0.011
rms angles (°)	1.62
Ramachandran favored (%)	97
Ramachandran outliers (%)	0.49
Clashscore	6.89
Average B-factor (Å^2)	37.6
macromolecules	37.8
ligands	31.8
solvent	33.5

*Statistics for the highest-resolution shell are shown in parentheses.

Table S3 - Principal Components Analysis of Fe-protein conformation

set	structure	ref FeP subunit	EV1 (Å)	EV2 (Å)	EV3 (Å)	hinge angle (°)
1	1N2C/alf	E	5.93	0.84	-0.43	0
2	1M34/alf	E	5.73	0.81	-0.39	0.3
3	1M34/alf	G	5.65	0.88	-0.33	0.4
4	1M34/alf	M	5.62	0.83	-0.35	0.5
5	1M34/alf	O	5.69	0.91	-0.41	0.5
6	pcp	E	2.91	-0.64	2.02	10.8
7	pcp	G	2.14	-1.40	0.88	10.9
8	pcpadp	E	2.46	-1.17	0.91	11.0
9	pcpadp	G	1.96	-1.40	0.76	11.1
10	2AFI/adp	E	-0.84	-1.19	-1.49	17.1
11	2AFI/adp	G	-1.84	-0.76	-0.17	19.8
12	2AFI/adp	M	-3.15	1.88	-0.32	26.4
13	2AFI/adp	O	-2.12	-1.87	-0.73	20.8
14	1FP6/adp	A	-1.34	-1.54	-0.97	18.4
15	2NIP/wt	A	-3.39	0.05	-0.81	25.1
16	1G5P/wt	A	-3.39	0.04	-0.81	25.1
17	2AFH/nf	E	-4.58	-0.12	0.29	27.9
18	1M1Y/xlink	E	-4.48	0.90	0.47	
19	1M1Y/xlink	G	-4.52	1.02	0.46	
20	1M1Y/xlink	M	-4.12	1.00	0.68	
21	1M1Y/xlink	O	-4.32	0.92	0.74	

EV1, EV2 and EV3 are the components along the top 3 eigenvectors calculated from the PCA of Fe-protein conformations. The rotation axes corresponding to EV1 and EV2 are depicted in text Figure 3. EV1 corresponds to a hinge bending motion as the two Fe-protein subunits qualitatively move either towards or away from the dimer interface, while EV2 corresponds to a twisting motion as the Fe-protein subunits move along the interface. For this calculation, a translation of 2 Å along an eigenvector corresponds to ~5° rotation about the corresponding rotation axis.

Coordinate sets are identified by PDB ID; sets 1-13 and 17-21 are Fe-proteins complexed to MoFe-protein (alf = *alf*-complex;[6,16] pcp = *pcp*-complex and pcpadp = *pcp/adp*-complex (this work); adp = *adp*-complex[2]; 1FP6 = free Av2 complexed to ADP;[17] wt = wildtype Av2;[18,19] nf = nucleotide free-Av2 complexed to Av1;[2] and xlink = Av2 chemically crosslinked to Av1[6])

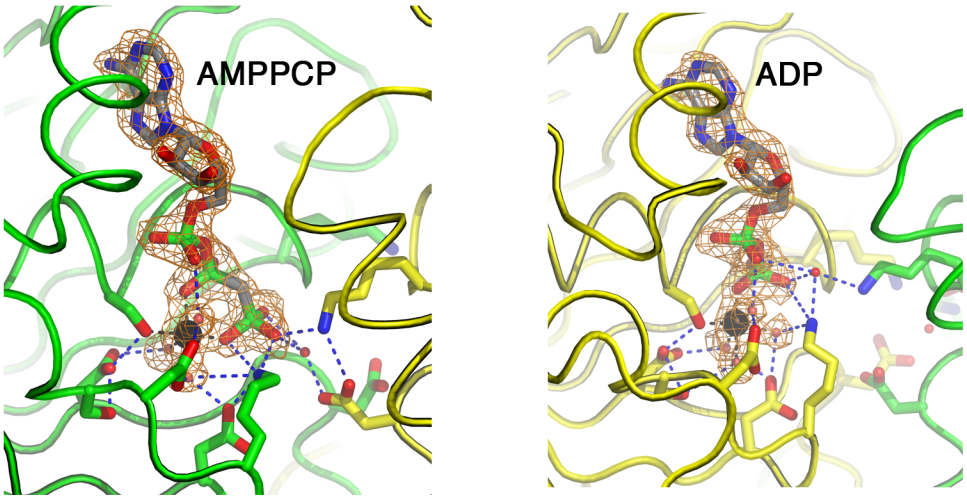

Figure S1. Simulated-annealing, omit F_o–F_c electron densities of Mg.AMPPCP and Mg.ADP nucleotides contoured at 4.0 σ.

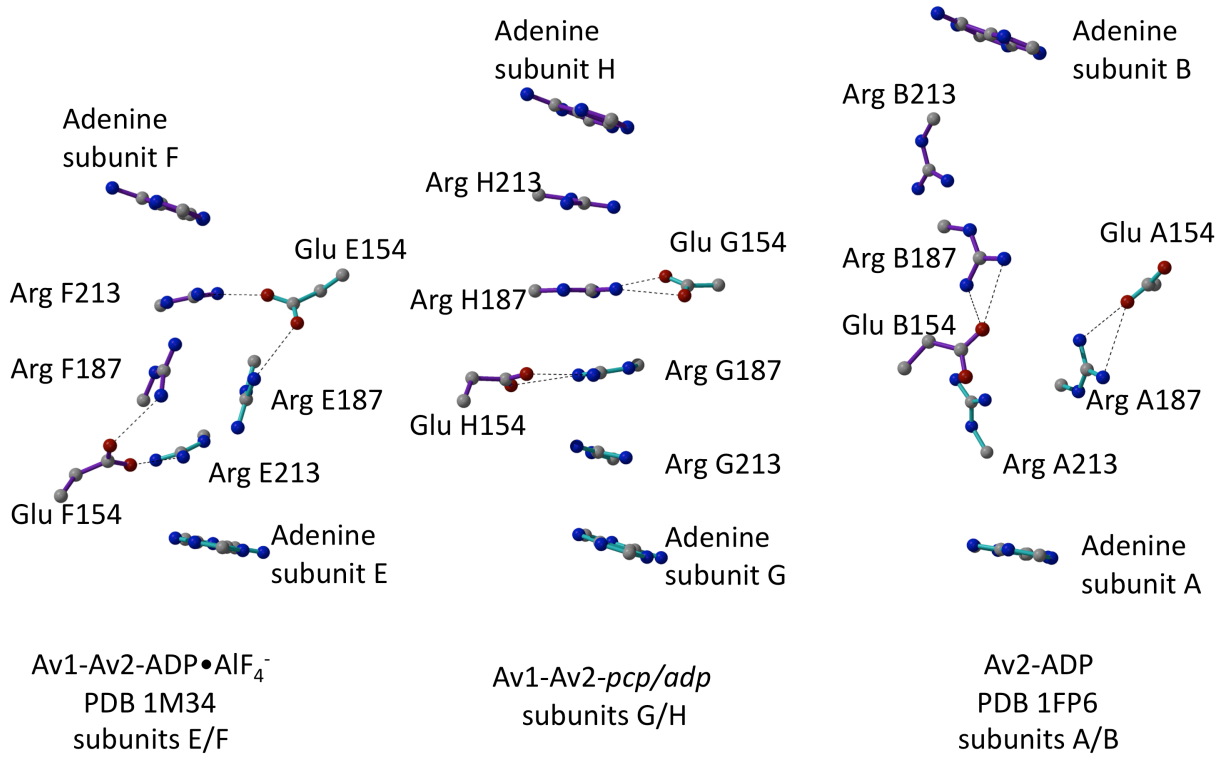

Figure S2. Glu154, Arg187 and Arg213 are observed to form alternative and mutually exclusive sets of inter- and intra-subunit salt bridges[16,18] that could serve to differentially stabilize the Fe-protein subunits in defined conformations with distinct subunit-subunit orientations. Glu154 exhibits three distinct sets of interactions, forming intersubunit salt bridges to Arg213 in the *alf*-complex, to Arg187 in the *pcp*- and *pcp/adp*-complexes, and an intrasubunit saltbridge to Arg 187 in the ADP state. An intriguing aspect of these interactions are the presence of various types of arginine stacking interactions, including the prominent Adenine-Arg213-Arg187-Arg187-Arg213-Adenine stack observed in the *pcp/adp*-complex structure; while stacking of guanidinium groups would be expected to be electrostatically unfavorable, several recent studies have highlighted the frequency of their occurrence, suggesting that the overall energetics are stabilizing.[20,21]

Supplementary Information - References

(1) Wolle, D.; Kim, C.; Dean, D.; Howard, J. B. *J. Biol. Chem.* **1992**, *267*, 3667.

(2) Tezcan, F. A.; Kaiser, J. T.; Mustafi, D.; Walton, M. Y.; Howard, J. B.; Rees, D. C. *Science* **2005**, *309*, 1377.

(3) Kabsch, W. *Acta Crystallogr.* **2010**, *D66*, 125.

(4) Evans, P. R.; Murshudov, G. N. *Acta Crystallogr.* **2013**, *D69*, 1204.

(5) Vagin, A.; Teplyakov, A. *J. Appl. Crystallogr.* **1997**, *30*, 1022.

(6) Schmid, B.; Einsle, O.; Chiu, H.-J.; Willing, A.; Yoshida, M.; Rees, D. C.; Howard, J. B. *Biochemistry* **2002**, *41*, 15557.

(7) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr.* **1998**, *D54*, 905.

(8) Turk, D. *Acta Crystallogr.* **2013**, *D69*, 1342.

(9) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallogr.* **2010**, *D66*, 213.

(10) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr.* **2010**, *D66*, 486.

(11) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr.* **2010**, *D66*, 12.

(12) DeLano, W. L.; DeLano Scientific: Palo Alto, CA, 2002.

(13) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct., Gen.* **1993**, *17*, 412.

(14) van Aalten, D. M. F.; Conn, D. A.; de Groot, B. L.; Berendsen, H. J. C.; Findlay, J. B. C.; Amadei, A. *Biophys. J.* **1997**, *73*, 2891.

(15) Kabsch, W. *Acta crystallogr.* **1976**, *A32*, 922.

(16) Schindelin, H.; Kisker, C.; Schlessman, J. L.; Howard, J. B.; Rees, D. C. *Nature* **1997**, *387*, 370.

(17) Jang, S. B.; Seefeldt, L. C.; Peters, J. W. *Biochemistry* **2000**, *39*, 14745.

(18) Schlessman, J. L.; Woo, D.; Joshua-Tor, L.; Howard, J. B.; Rees, D. C. *J. Mol. Biol.* **1998**, *280*, 669.

(19) Strop, P.; Takahara, P. M.; Chiu, H.-J.; Hayley, C.; Angove, C.; Burgess, B. K.; Rees, D. C. *Biochemistry* **2001**, *40*, 651.

(20) Neves, M. A. C.; Yeager, M.; Abagyan, R. *J. Phys. Chem. B* **2012**, *116*, 7006.

(21) Lee, J.; Seok, C. *Phys. Chem. Chem. Phys.* **2013**, *15*, 5844.